Präsenzaufgabe 2.1

Doubling trick : Aus der Vorlesung ist bekannt, dass die maximale kumulative Reue R_n^* des "exponentially weighted average forecasters" für alle $\eta > 0$ abgeschätzt werden kann (wenn $\ell : \mathcal{D} \times \mathcal{Y} \to [0,1]$ konvex in erster Komponenten ist) durch:

$$R_n^* \leq \frac{\ln N}{\eta} + \frac{\eta}{8} n \,,$$

wobei der Horizont n vorab bekannt sein muss.

Wir teilen nun die Zeit für ein $a > 1$ in Phasen $r = 0, 1, 2, \ldots$ der Länge $\lfloor a^r \rfloor$ (Runden $\lfloor a^r \rfloor, \ldots, \lfloor a^{r+1} - 1 \rfloor$) und verwenden während Phase r den Parameter $\eta_r = \sqrt{8 \ln N / a^r}$.

a) Zeige, dass für $a = 2$ die kumulierte Reue für einen solchen Forecaster wie folgt beschränkt werden kann:
$$R_n^* \leq \frac{\sqrt{2}}{\sqrt{2} - 1} \sqrt{\frac{n}{2} \ln N} \,.$$
b) Welche Wahl von $a > 1$ ist asymptotisch optimal? Wie verhält sich diese asymptotische Schranke zur optimalen Schranke aus der Vorlesung, die unter Kenntnis des Zeithorizonts abgeleitet wurde?

Präsenzaufgabe 2.2

Aus der Vorlesung ist bekannt, dass der kumulative Verlust des "exponentially weighted average forecaster" für alle $\eta > 0$ wie folgt nach oben abgeschätzt werden kann (wenn $\ell : \mathcal{D} \times \mathcal{Y} \to [0,1]$ konvex in erster Komponenten ist):

$$\hat{L}_n \leq \frac{\eta L_n^* + \ln N}{1 - e^{-\eta}} \,.$$

Zeige, dass bei der Wahl von $\eta = \ln \left(1 + \sqrt{(2 \ln N)/L_n^*} \right)$ der kumulative Verlust R_n^* wie folgt beschränkt werden kann, wenn wir annehmen, dass $L_n^* > 0$ vorab bekannt ist:

$$R_n^* \leq \sqrt{2 L_n^* \ln N} + \ln N \,.$$

Hinweis : Nutze die Abschätzung $\eta \leq \sinh \eta$.

Aufgabe 2.1 (4 Punkte)

Wir wählen $\mathcal{D} = [0,1]$, $\mathcal{Y} = \{0,1\}$ und die Verlustfunktion $\ell(x,y) = |x - y|$. Zeige, dass unter diesen Annahmen die Strategie des gradientenbasierten Forecasters zu denselben Vorhersagen $\hat{p}_t$ führt wie die des "exponentially weighted average forecaster".

Bemerkung: Erweitere die Strategie der gradientenbasierten Vorhersage für die Definitionslücke von $\nabla \ell(\cdot, y)$ an der Stelle $x = y$ geeignet.

Aufgabe 2.2 (4 Punkte)

Doubling trick (II) : Für die Schranke in Präsenzaufgabe 2.2 müssen wir L_n^* vorab kennen. Verwende den "doubling trick" um die folgende Schranke (unabhängig von L_n^*) zu beweisen:

$$R_n^* \leq 2\sqrt{L_n^* \ln N} + (\log L_n^* + 1) \ln N \, .$$

Aufgabe* 2.3 (4 Punkte)

Die Resultate aus der Vorlesung, wurden für eine uniforme Verlustfunktion $\ell : \mathcal{D} \times \mathcal{Y} \to \mathbb{R}$ gezeigt. Zeige, dass die Schranken auch für parametrisierte Familien von Verlusten $(\ell_t)_{t=1,2,\ldots}$ gelten, die in jeder Runde den Verlust nach ℓ_t bemessen.

Führe die Analyse dabei auf den Fall der uniformen Verlustfunktionen zurück. *Hinweis:* Nutze eine Erweiterung des $\mathcal{D}$-Raumes um die Vorgeschichte geignet zu kodieren.